Interactive Trajectory Modification and Generation with FPCA

Gabriel Jarry, Almoctar Hassoumi, Daniel Delahaye, Christophe Hurter

Interactive Trajectory Modification and Generation with FPCA

Gabriel Jarry, Almoctar Hassoumi, Daniel Delahaye, and Christophe Hurter
Ecole Nationale de l'Aviation Civile, Université de Toulouse
7 Avenue Edouard Belin, 31400 Toulouse, France
{gabriel.jarry, almoctar.hassoumi, daniel.delahaye, christophe.hurter}@enac.fr

***Abstract*—Moving object analysis is a constantly growing field with numerous concrete applications in terms of traffic understanding, prediction and simulation. While many algorithms and analytic processes exist, there are still areas of investigation with novel trajectory analysis methods. As such, the geometric information analyses data with respect to its statistical distribution along extracted dimensions. This opens new ways of gaining a better understanding of large and complex trajectory data sets while providing flexible data manipulations. In this paper, we report our investigations with the development of an interactive methodology based on the geometric information analytic process where users can analyze trajectories sets, cluster and deform them maintaining the actual statistical properties of the investigated trajectories. As a contribution, this paper shows how geometric information can provide novel support for trajectory analyses taking into account the statistical properties of the investigated clusters. We also provide recommendations of good usage of such techniques with actual examples validated by a a domain expert of air traffic flow analysis.**

***Index Terms*—Geographic/Geospatial Visualization, Data Aggregation, Data Cleaning, Data Clustering, Data Transformation and Representation, Data Editing, Manipulation and Deformation, Multidimensional Data, Geometry-based Techniques**

INTRODUCTION

Our society has entered a data-driven era, in which not only enormous amounts of data are being generated every day, but also growing expectations are placed on their analysis [1]. Trajectory data (i.e. flows of cars, airplanes or people) are collected every day and analyzing these massive and complex data sets is essential to making new discoveries and creating benefits for people. Processing such data is a challenging task due to their intrinsic, time-dependent nature. While machine learning heralds a solution to address the issues of big data and efficient knowledge extraction, alternatives do exist where humans play a central role with the usage of interactive visualization systems [2].

In this regard, this paper investigates a novel analytic method for trajectory processing using information geometry [3]. While general trajectory analysis relies on distance and time algorithms, information geometry uses differential geometry and probability theory [4]. Such analytic tools capture the intrinsic statistical properties of the investigated trajectories. Previous work [5] showed its potential to support visual simplification and visual flow modeling. Geometry information deserves further investigation which goes beyond its usage for visualization purposes.

Considering trajectory input data as a set of $\mathbb{R}^d$ curves, the standard multivariate statistical representation of a set of curves γ would be a set of d-dimensional samples [6]–[8]. However, this representation may not capture all relevant curve characteristics - e.g. its shape or smoothness. Functional data analysis [9] enables a better representation of multivariate data functions like curves. A curve is then modeled as a point in an infinite-dimensional space, usually the L^2 space of square-integrable functions [9]. Geometry information can then be used to obtain a finite representation of the data by means of Functional Principal Component Analysis (FPCA). This tool captures the data variability around the mean curve while estimating the Karhunen-loève expansion [9]. In other words, FPCA yields a finite basis describing the main variability modes contained in the data. Learning the distribution of the data on this basis enables two powerful applications: the generation of new samples with the same behavior, and the creation of samples with a user-deformed mean consistent with the collected data.

This paper applies geometry information for analytic purposes and proposes an analytic pipeline to support trajectory processing. This pipeline handles trajectory clustering, data cleaning, flow simplification, flow generation and flow transformation. This methodology was built with the help of air traffic experts to ensure the accuracy of the processed information. This paper's contributions rely on the analytic pipeline and its guidelines to leverage trajectory analysis with geometry information tools.

The article is structured as follows: Section I presents related works on existing trajectory processing algorithms. Section II lays the mathematical foundations for trajectory analysis limited to flow understanding and management. The following section gives the basis of the information geometry. Next, we detail the pipeline followed by its use cases. Next, we discuss this paper with an extract recommendation for good usage of the tools. Finally, we conclude the paper with possible work extension.

I. RELATED WORK

There is abundant literature concerning the analysis of moving object trajectories. Even if it is a well-explored topic, it remains a popular area of research where geometry information has barely been used [5]. This paper fills this gap with a thorough usage of geometry information-based algorithms for trail set analysis and deformation. This section presents the main research challenges encountered in the fields of trajectory analysis, trajectory deformation and interactive exploration of trajectories.

A. Functional modeling

When manipulating objects that have a functional nature or are raised from a functional model, it is advised to preserve this model by using appropriate tools. Functional Data Analysis (FDA), [9] is a tool that aims to precisely preserve the functional nature of data by expanding it into an appropriate finite functional basis. The input object is transformed into a coefficient vector, which can then be used in a multivariate framework. This enables the use of traditional multivariate statistics but with the insurance of keeping the functional behavior of the underlying objects. The choice of the basis is important since an appropriate basis choice can better capture some data features such as smoothness. Nevertheless, the main applications of FDA are real-valued functions such as spectrometric data [10] or weather data [11]. Nevertheless, there are few applications on vector-valued functions, such as the 2D or 3D curves considered in [12], and [13], [14] or trail set brushing [15].

B. Trajectory clustering, simplification, and generation

Trajectory analysis often relies on clustering algorithms. Clustering can be performed on the geographical space [16] with density maps [17], with pattern similarities [18] or with time clustering [19]. It is also possible to define distances between trajectories to enable clustering [20], or to use dimensional reduction processes [12]. Using the Functional Principal Component Analysis for trajectory clustering has barely been investigated yet, which makes this study a precursor in the area.

New methods such as Generative Adversarial Networks allow to generate trajectories. A recent publication [21] proposes the use of GANs for the generation of aircraft trajectories and the detection of atypical approaches.

C. Trajectory Exploration tools

Exploring, analyzing and visualizing temporal data such as trajectories has a long history. Time series analysis [22] helps the extraction of relevant information. Frameworks [23] are available to gain a better understanding of such complex time-varying data sets thanks to aggregation techniques [24]. A recent visualization framework has been provided to structure efficient temporal data representations [25].

Many interactive tools and systems for trail-set exploration and manipulation exist. Selection boxes help to filter objects of interest [26] [27], particle systems help to understand flow directions [28]. More recently, image based techniques [29] have been applied for trajectory analysis [30]. Boolean operation can be performed to combine selections of trajectories [31] on a 2D screen or in virtual reality [32]. Overall, no previous system used the FPCA tools in a unified framework for trajectory analysis and this paper provides the first of the kind.

II. Mathematical foundations

This section provides the mathematical foundations to understand the Functional Data Analysis process. Section II-A underlines the representation of discrete trajectories in a function space. Section II-B explains how principal curves help to represent a set of trajectories. Section III-A shows how to generate a new distribution using these principal coefficients. Finally, Section IV illustrates how to modify curves thanks to controlled deformations.

A. Curve functional modeling

Functional Data Analysis considers curves as objects in an infinite dimensional space. This enables certain curve behaviours such as their shape or smoothness to be taken into account. To retrieve the functional model from discrete data, curves must be reconstructed in a dedicated functional space. It is mandatory that curves have two continuous first derivatives and thus belong to the L^2 space of square-integrable functions. Before applying functional decomposition, curves must belong to this space, so called Sobolev [33].

$$\mathbb{W}^2 = \left\{ f \in C^1([0,1],\mathbb{R}), f' \text{ abs. cont.}, \int_0^1 f(x)^2 + f''(x)^2 dx < +\infty \right\}. \tag{1}$$

To obtain a functional representation of the discrete curves, the choice of a cubic spline kernel K is made based on [5]. A set of curves $P = \{\gamma_i\}$ is then represented by a matrix A where each row a_i represents γ_i in terms of spline coefficients.

B. Functional Principal Component Analysis modeling

Let $C = \{\gamma_1, \ldots, \gamma_N\} \subset P$ be a set of N curves. The Functional Principal Component Analysis (FPCA) process consists in modeling C with its mean curve $\bar{\gamma}$ and the variance around it. A classic hypothesis is that C comes from an underlying hidden stochastic process $\gamma : \Omega \times [0,1]$, where Ω is the probability space of all possible outcomes. The empirical covariance estimator $\hat{H}$ enables the capturing of the variability of C around its mean $\bar{\gamma}$ by using the Karhunen-Loève expansion [34]:

$$\Gamma(t,\omega) = \bar{\gamma} + \sum_{j=1}^{+\infty} b_j(\omega)\phi_j(t) \tag{2}$$

where b_j are real-valued random variables called principal component scores. ϕ_j are the (vector-valued) eigenfunctions of the covariance operator with eigenvalues λ_j. For the discrete implementation of such functional decomposition see [5]. With this model and knowing the mean curve $\bar{\gamma}$ and the principal component functions ϕ_j, a group of curves can be described and reconstructed (Inverse FPCA) with the matrix of the principal component score b_j of each curve. Usually, a finite vector (fixed dimension d) of b_j scores is selected such that the explained variance is more than a defined percentile.

To sum up, each trajectory can be represented through the FPCA process by the *Mean* plus the sum of the *Principal Component Functions* weighted by the *Principal Component Score*. The Inverse FPCA (IFPCA) process consists in reconstructing the trajectory from the *Principal Component Scores* knowing the *Mean* and the *Principal Component Functions*

III. Tools

This section is divided into two parts. Section III-A explains the curve generation process, and Section III-B presents the clustering task.

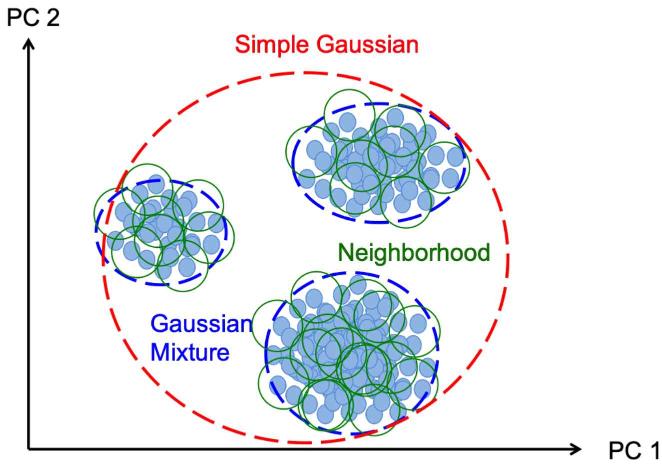

Fig. 1: This figure illustrates the different generation processes for the principal component score distribution. In red, a Simple Gaussian model is illustrated. In blue, the Mixture Gaussian model is represented. Finally, in green, the neighbourhood model is shown

A. Curve generation

In [5], Hurter et al. generated curves with a random selection of principal coefficient scores with a centered independent simple Gaussian distribution hypothesis. Usually, coefficients are not simply Gaussian. Consequently, curves generated with this model do not present realistic behavior. Two alternatives are proposed in the following.

1) Neighborhood generation: To ensure the generated curves are sufficiently realistic, a neighborhood generation was developed in this study. The curve regeneration process needs to take into account an important number of principal coefficients (typically more than 60% of the total principal components). The process is as follows: First, each principal coefficient of the dimension variance is computed using a Gaussian centered model. Then, a curve is randomly selected and its principal component scores are kept. Finally, a new score is randomly generated in the neighborhood of the selected sample. The range of the neighborhood is defined with the variance among each dimension. In addition, the user is able to tune a α coefficient between 0 and 1 that is multiplied to the range of the neighborhood. This coefficient enables the user to modify the similarity between the original and the generated trajectory.

2) Multivariate Gaussian mixture model generation: An alternative to the neighborhood generation model consists in applying a multivariate Gaussian Mixture model, i.e. an EM algorithm [35], on the principal component scores that concentrate more than a user-defined percentage of the explained variance. This process does not assume the independence of the principal component scores and enables a richer representation with a Gaussian Mixture instead of a simple Gaussian Distribution. With this generation, it is usually more difficult or even impossible to properly estimate the distribution for a large number of components. This is the well known problem referred to as the curse of dimensionality [36]. In high dimensional space, the volume of space increases rapidly and samples are usually isolated. The choice was made in this study to estimate only the distribution of the first components that explain most of the variance with the dependence hypothesis. The last components, which mostly correspond to the noise, are then assumed to be independent.

Figure 1 illustrates the three models for curve generation. The red circle corresponds to the Simple Gaussian model, the blue ones to the Gaussian Mixture Model, and the green ones to the neighborhood model.

B. Clustering

1) Clustering for regeneration: Clustering is a very important initial step before applying the FPCA process. To be efficient, FPCA must operate on clusters with representative mean curves. A two step clustering process was derived for this study. A cutting down clustering, which aims at reducing large data-sets, is applied in a first step. For example, one can use a k-means clustering (or other simple literature algorithms) on arrival or departure trajectory locations. For the study of aircraft landing trajectories, the initial clustering is here done on the destination runways. The second step is to apply a refinement clustering based on the FPCA decomposition score. Displaying first coefficient dimension, the user is able to apply another clustering algorithm in order to group together similar trajectories. The choice of the Esperance-Maximization (EM) algorithm [35] is made here but other algorithms such as k-mean [37], or HDBSCAN [38] are also applicable. The choice of the algorithm and/or the number of clusters should be guided by the visualization of the FPCA score and by expert knowledge of the investigated data-set. In addition, the user is able to select the number of dimensions of the principal component score to use for the clustering and visualize the clustering result on the trajectory to decide which clustering method produces the most representative clusters (i.e. mean curve dissimilarity).

2) Clustering for Classification: The distribution of the principal component score can be used to cluster data. Indeed, the finite dimension representation enables the computation of distance. Besides, the euclidean norm of the principal component score is equal to the L2-norm in the Sobolev Space [9]. In a situation where the behavior of the group of trajectories to classify is known, this knowledge can be used to define a classification process using unsupervised learning techniques. First, trajectories are decomposed using the FPCA process. Then, the HDBSCAN [38] clustering algorithm is applied to all the trajectories principal component scores. Since the FPCA process clusters together similar data, it means that similar trajectories will be grouped together. HDBSCAN is really highly efficient in determining density-based clusters with irregular shapes, i.e. clusters that are generated from the same distribution with no assumptions on the type of distribution. In addition, the HDBSCAN algorithm gives the probability of being in a cluster. Knowing the behavior of the group to be detected, it is possible to identify to which cluster it corresponds. Finally, the user defines a probability value above which the trajectory is attributed to a cluster. This enables the user to choose the characteristics of their classification algorithm in terms of accuracy or specificity.

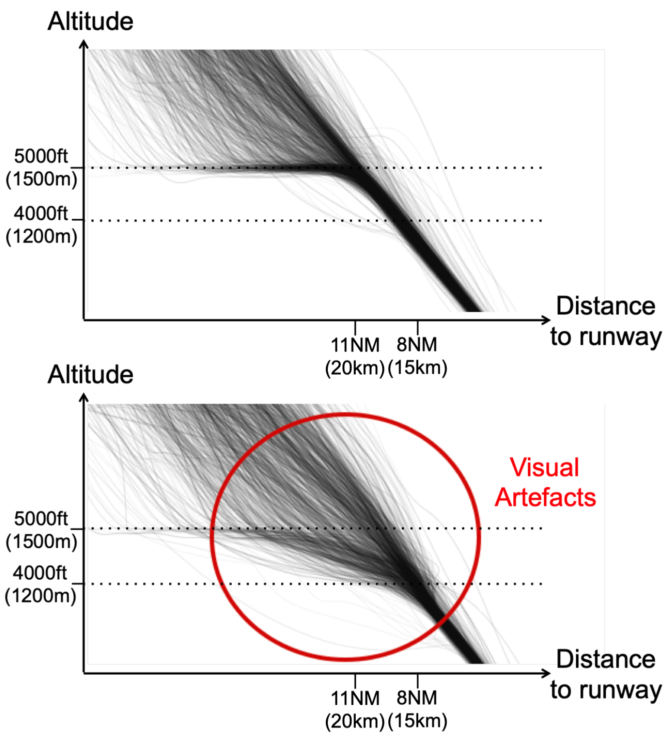

Fig. 2: This figure illustrates examples of good and poor usage of curve modifications for aircraft approach trajectory altitude profile. At the top, both the mean and the principal components were modified. At the bottom, only the mean was modified. The behavior observed while modifying only the mean curve presents artifacts. Level-off flight was expected, but the trajectories present descent phases. These behaviors are not nominal and underline that the process was not executed properly.

IV. Curve shape modification

While simple trajectory deformations can be performed with Cartesian dimensions, it become more complex with additional data dimensions such as altitude. Furthermore, deformation becomes cumbersome when it has to be applied to many trajectories. FPCA can help solely with the deformation of the cluster mean curve and its principal components to modify every trajectory of the investigated cluster. Hurter et al. [5] only modified the mean curve to perform trajectory modifications, which leads to many visual artifacts. Indeed, the mean curve modification is not sufficient, the principal components also have to be modified to correctly model the temporal behavior which was embedded in the undistorted original FPCA model. Modifying the trajectory behavior implies being sure that the principal components, and therefore their underlying variation on the mean curve behaviour, are applied at the right time-stamp. Modifying the mean curve without insuring that the role of the principal components was not modified, resulted in most cases, with aberrant curve behaviors.

In figure 2 good and poor usage of curve modifications are illustrated. This shows aircraft vertical profiles (altitude function of the distance) modifications, where the landing procedure was increased in altitude (1000ft higher for noise sustainability issues). This use case will be further detailed in section VI-B1. The top figure 2, shows the result of the solely mean curve modification. In this case, the principal components are no longer aligned with the mean curve and result in an unrealistic trajectory with artifacts around the level-off flight (red circle in figure 2). The bottom figure 2 shows more realistic results where both the mean and the principal components were modified.

The mean curve modification and its principal component modifications are not an easy task. First, a curve registration is needed to align curve landmarks. Then, the key idea is to apply the same temporal modifications to both the mean and the principal components. By doing so, we ensure that the variation induced by the principal component function is correctly located.

Curves translation: The translation operator of a curve γ is defined as v the translation vector for any $t \in [0,1]$ as :

$$\text{Translation}_{\gamma}(v)(t) = \gamma(t) + v \tag{3}$$

This is the sole operator that can be applied only to the mean curve since it does not affect landmark time position.

Curves 2D rotation: The 2D rotation of a curve $\gamma = (\gamma_x, \gamma_y)$ at time t_1 with angle θ for any $t \in [0,1]$ as is defined as:

$$\text{Rot}_{\gamma}(t_1, \theta) = \begin{cases} \gamma(t), & \text{if } t < t_1 \\ \gamma(t_1) + \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} (\gamma(t) - \gamma(t_1)) & \text{else} \end{cases}$$

(4)

Temporal compression and dilatation: Temporal compression/dilatation operator of a curve γ between t_1 and t_2 with the compression coefficient $\alpha \in \mathbb{R}^+$ with $t_{CD} = \frac{t_2 - (1-\alpha)\cdot t_1}{\alpha}$, the temporal compression/dilatation is defined for any $t \in [0,1]$ as :

$$\text{TCD}_{\gamma}(t_1, t_2, \alpha)(t) = \begin{cases} \gamma(t), & \text{if } t < t_1 \\ \gamma(\alpha \cdot t + (1-\alpha) \cdot t_1), & \text{if } t_1 \leq t < t_{CD} \\ \gamma(t - (t_{CD} - t_2)), & \text{if } t \geq t_{CD} \end{cases} \tag{5}$$

Curve temporal cut or extension: The cutting or extending operator of a curve γ at t_1 with width $\delta_t \in [-t_1, 1-t_1]$, for any $t \in [0, 1-\delta_t]$ is defined as :

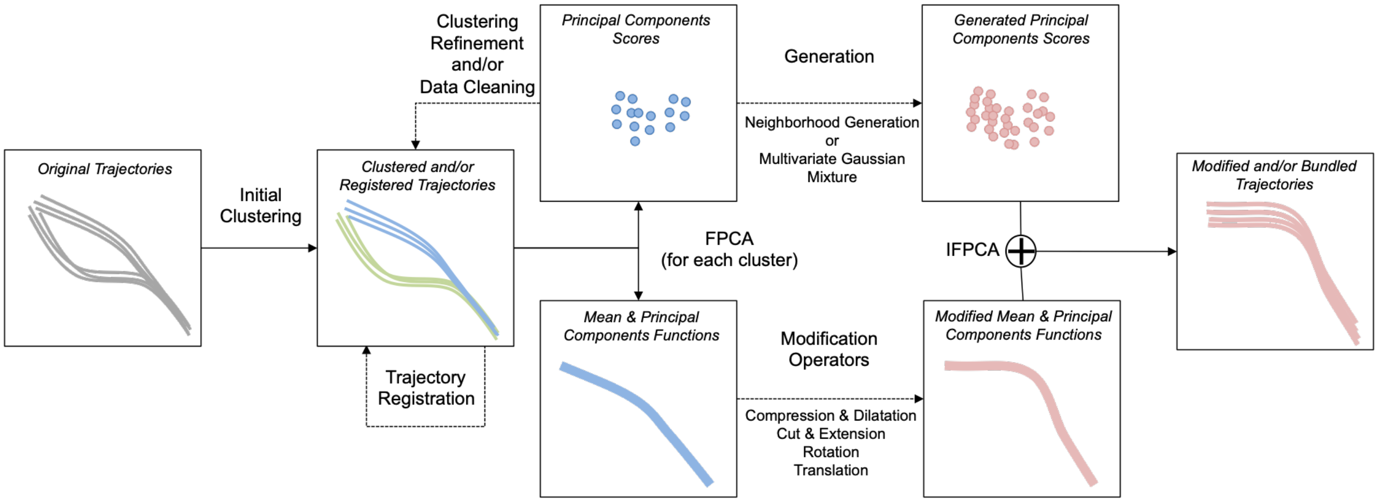

Fig. 3: Illustration of our pipeline. On the left, the input curve data-set passes through an initial clustering step. After this, trajectory registration based on landmarks is applied. Then, the FPCA process is computed for the first time for each cluster. This FPCA process gives two pieces/elements of information: the principal component scores (top figure) and the mean curve with the principal component functions (bottom figure). The principal component scores are used in two situations: First for clustering refinement and data cleaning, second for the generation process. Generation consists in estimating the principal component score distribution and in generating new samples following the estimated distribution. During the modification operations, the mean and the principal component functions are processed. It consists in applying modification operators (rotation, translation, dilatation) in order to obtain the desired distortion. Finally, the Inverse FPCA process enables the trajectory to be reconstructed with the new distribution (i.e. increase or decrease in trajectory number) and behavior.

$$\mathrm{CE}_\gamma(t_1,\delta_t)(t) = \begin{cases} \gamma(t), & \text{if } t < t_1 \\ \gamma(t+\delta_t)+\gamma(t_1)-\gamma(t_1+\delta_t), & \text{if } t_1 \le t < 1-\delta_t \end{cases} \tag{6}$$

In addition, this operator modifies the definition interval of the curve. A good use consists in applying the TCD (Eqn. 5) operator between 0 and $1-\delta_t$ with compression coefficient $\alpha = 1-\delta_t$.

Smoothness: The three last operators (Eqn. 6, Eqn. 5, Eqn. 4, only insure the continuity but do not ensure the smoothness of the obtained curve. Nevertheless, it may be restored for an operational or visual purpose by using an additional filtering algorithm (Laplacian filtering or other).

V. Pipeline

The figure 3 shows the pipeline with a trajectory data-set as input data. The first step performs an initial clustering to reduce the data-set size into clusters with similar trajectories. This initial clustering is data-set dependant. For instance, with aircraft trajectories, it can be performed on departure or arrival airport. The FPCA process can then be applied to each cluster to compute the mean curve, the principal component functions and the principal component scores.

Then, a clustering refinement step can be computed based on the principal component scores as previously explained. User input is needed at this clustering step: with the suggested EM algorithm, the user has to define the number of clusters and the number of principal components to use.

The following step shows different possible trajectory processing. The Inverse FPCA produces size varying trajectory with respect to their shape and statistical properties. Two different types of trajectory generation are available (III-A). The neighborhood process produces trajectories close to the original one (tuning the variability around it), the Gaussian mixture allows more variability. Finally, trajectory deformations can be applied to distort the final result. Thanks to modifications operators (IV) applied on the mean and on the principal component functions, trajectory shape can be adjusted. These final data processing techniques (curves generation, deformation and simplification) can be combined to to adjust end user final results.

VI. Use cases

This section is divided into two parts. Each part illustrates a specific feature of the pipeline (Figure 3). The first part illustrates the clustering process, the second one the trajectory modification operator through concrete examples.

A. Clustering and Classification

In this section we will apply the unsupervised classification process defined in section III-B2 to the identification of landing procedure at Bordeaux Merignac airport (one of the major airports in France). When landing on runway 05, aircraft follow four kinds of trajectory (RNAV,Visual-RNAV, VOR, Conv). The RNAV approaches, are GNSS paths. They are very characteristic since they follow a path from defined way-points (geographical points on a map). It also means that this type of approach will be very similar in the FPCA space. We recorded 2597 suitable landing sequences (one record every 4 seconds, 995963 points in total) during a three month period in 2018 (summer time). We fixed the probability threshold of each cluster in order to minimize the false positive samples. Indeed, we know the behaviour of these approach trajectories, so we fixed the threshold such that all the detected trajectories correspond to this behavior.

Figure 4 shows the result of the clustering algorithm where classes of landing sequence are clearly separated within four clusters. Figure 4 bottom shows the first four principal components with the identified cluster.

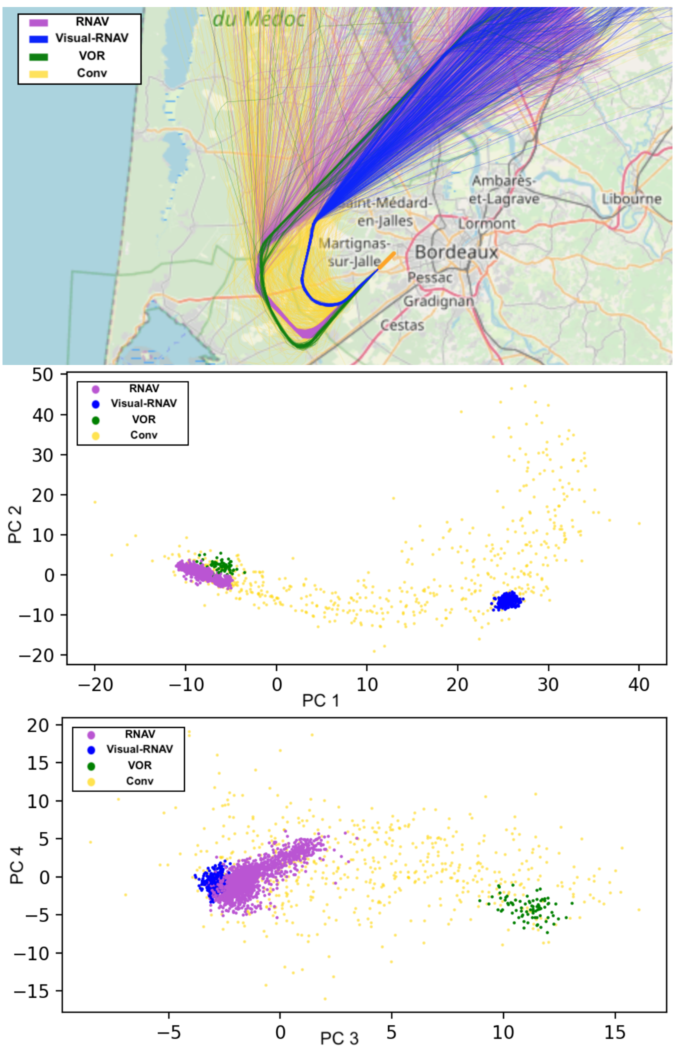

Fig. 4: This figure illustrates the clustering task of the landing trajectory at Bordeaux airport. On the top image is illustrated the result of the classification of the Bordeaux airport approaches from the 2017 data records. At the bottom, the first four principal component score distributions are represented.

B. *Curve modification and generation*

1) Interception Altitude modification: After The Grenelle de l'environnement 2018 annual meeting to discuss sustainability issues, landing procedures at Charles de Gaulle airports were raised by 1000ft (around 300m) to reduce noise emission. In this regard, we worked in collaboration with the Environmental Office of the French Civil Aviation Authority. In this section, we report the simulation results where we processed traffic before the rise and modified them with the Grenelle 300 meter rise. We then computed the resulting noise emission and compared it with the actual trajectories after the rise. This comparison provides a good assessment of the accuracy of the trajectory generation and modification pipeline.

The process is the following. First, curves were registered with their landmarks defined along the longitudinal turning points of the curve. Second, the FPCA decomposition and a clustering refinement of the trajectories with the EM algorithm [35] were applied.

Then, for each cluster, curves were modified to follow the 300 meter pull-up: extension operator from 1000ft to 2000ft with t_1, the time shift at 1000ft (300m), t_2 the time shift at 2000ft (600m), and $\delta_t = t_2 - t_1$.

Figure reffig:noiseSimu top illustrates the noise level for the real traffic (after the altitude rise), and at the bottom, the noise generated from the pipeline with the modified traffic. The noise indicator is the NA62 indicator which is computed over one day of traffic. This indicator is mainly used by the environmental office. It corresponds to the number of aircraft emitting noise above 62dB during the period. The area for 5 to 20 events above this threshold is represented here.

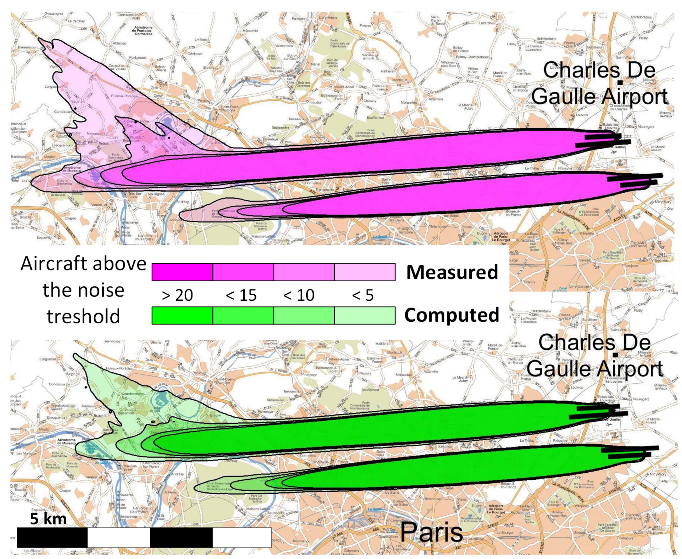

Fig. 5: At the top, the noise NA62 indicator map of the real aircraft traffic (aircraft noise above 62dB). At the bottom, the same indicator map for the simulated traffic obtained with the pipeline. The modification consists in raising by 1000ft (300m) the level-off flight of landing aircraft before landing. This modification was applied following the Grenelle de l'environnement for noise reduction purposes.

The result shows that this noise computation is close to the actual recorded noise with main identical parts even if a few differences in terms of areas can be observed. The simulation covers 92% of the area because the real noise map is slightly more extended on the left side. This is due to the fact that in the real context, approaches tend to have a longer level-off flight before starting the final descent. Nevertheless, this shows that the pipeline can produce valuable simulated trajectories and can be used for realistic flow simulation.

2) New departure flow investigation: Before building a new aircraft departure air flow, it is valuable that it is simulated and its noise impact assessed. This study's pipeline assists in this matter. To test such a possibility, we considered a novel departure air flow at Nantes-Atlantic Airport (one of the major airports in France) using an existing flow at Bordeaux-Merignac Airport as a reference model. Such flow duplication is not straightforward since the original air flow (i.e. Bordeaux) has to be modified to follow mandatory way points at the destination airport (i.e. Nantes).

As a result, figure 6 represents at the topthe original flow of trajectories at Bordeaux and at the bottom the generated and distorted air flow with the pipeline after its modification to fit Nantes airport landing procedure. A rotation and translation were applied to align the runway approach with the runway in Nantes. The new procedure follows different aeronautical way-points and is illustrated in red in figure 6.

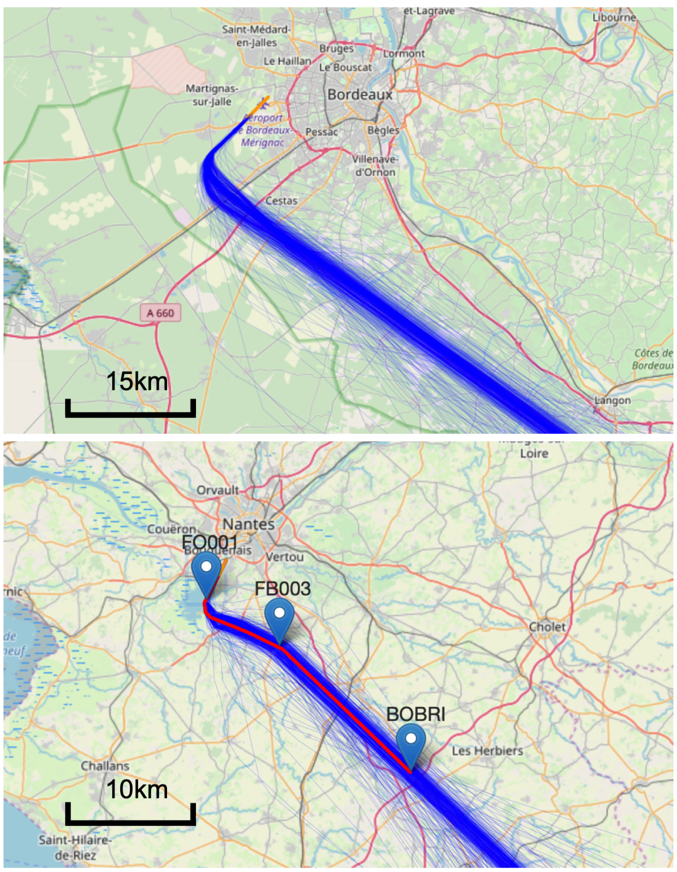

Fig. 6: This figure illustrates the departure procedure deformation. On top, the original trajectories at Bordeaux Airport are represented. At the bottom, the modified trajectories at Nantes Airport. In addition, the expected procedure following different aeronautical way points is shown in red.

VII. Discussion and FPCA good practice

This section discusses the paper outcomes (clustering, trajectory distortion, trajectory generation) and provides a summary of FPCA good practice.

The objective of clustering refinement is to compute cluster with consistent mean curve. As previously explained, the FPCA process is efficient when groups of curves have similar shapes to correctly capture their variability around a mean curve. In addition to the clustering refinement, it is also important to remove outliers which may impact the mean curve and potentially induce weak results in the next FPCA processing steps.

The system, derived from the pipeline, contains a set of specific tools for trajectory modifications. Modifying the mean curve without keeping the principal components aligned will generate artifacts. Users need to visually assess the modification results and fine tune the regenerated curves. Currently, the methodology has predefined modification presets, but in our future work, the user will be given the ability to choose the modifications and interact on the mean and on the principal components to directly see the effect on the regenerated curves.

For the generation process, the user has to select which kind of generation process they wish to apply and define how many curves to generate. The size of the original data-set can be adjusted while keeping a consistent distribution or extending the data-set in number in order to simulate traffic growth or decrease. The choices in the generation process and in the parameters are guided by the visualizations of the principal component score and the desired proximity in shape to the original trajectories. Besides, the user can also adapt the generation process with the visualization of the reconstructed trajectory and ensure that the generated curves have the shape expected.

As a summary, recommendations for efficient usage of FPCA tools for trajectory analysis are provided here:

- Trajectory registration: This initial step is mandatory to efficiently capture the variance around the mean curve of the considered clusters,
- Initial clustering: An initial clustering is mandatory to have a meaningful mean trajectory,
- trajectory deformation: trajectory deformation only operates with the mean curve deformation associated with the principal component function modifications,
- Trajectory generation: many possible methods exist to increase or decrease the number of trajectories. We proposed three methods taking into account the global, local and neighbor variance.

VIII. Conclusion

In this paper, we propose a new approach to analyze trajectories from a functional decomposition perspective for the underlying data-set. Thus, we developed a functionally based pipeline to support the following trajectory processing: clustering, trajectory deformation and trajectory generation. Thanks to the pipeline, trajectories can be clustered taking into account trajectory curvature and their variability around the mean curve. This provides another clustering tool which mainly considers trajectory shapes as a grouping parameter. Through concrete examples, an aircraft path deformation and the corresponding noise computation, we show that the pipeline can produce reliable solutions for trajectory simulation. The results seem to be relevant regarding operational metrics. Rather than processing every trajectory to deform it and make it compliant with new air traffic flow constraint, the pipeline enables the deformation of a single mean curve to produce an equivalent result. Furthermore, this pipeline is flexible, since the user can also increase or decrease the number of trajectories while keeping a coherent distribution around the mean curve.

While we show quantitative and accurate results with this pipeline, many improvements can be considered. Firstly, FPCA tools need some fine tuning and the underlying parameters require some prior knowledge in statistical tool manipulations. Secondly, the pipeline provides trajectory deformations applied to the mean curve and the principal component of the considered cluster. We currently provide simple transformation like rotation, stretching and bending. some additional work is needed to make this transformation applicable to any kind of trajectory.

References

[1] J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE Comput. Graph. Appl.*, vol. 26, no. 1, pp. 10–13, Jan. 2006. [Online]. Available: http://dx.doi.org/10.1109/MCG.2006.5

[2] J. Heer and S. Kandel, "Interactive analysis of big data," *XRDS*, vol. 19, no. 1, pp. 50–54, Sep. 2012. [Online]. Available: http://doi.acm.org/10.1145/2331042.2331058

[3] S.-i. Amari, *Information Geometry and Its Applications*, 1st ed. Springer Publishing Company, Incorporated, 2016.

[4] ——, *Differential-Geometrical Methods in Statistics*, 1st ed. Springer-Verlag New York, 1985.

[5] C. Hurter, S. Puechmorel, F. Nicol, and A. Telea, "Functional decomposition for bundled simplification of trail sets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 500–510, Jan 2018.

[6] H. Mark and J. Workman, *Statistics in Spectroscopy*. Elsevier Science, 2003. [Online]. Available: https://books.google.fr/books?id=qTZIYUrSA_oC

[7] R. Prado and M. West, *Time Series: Modeling, Computation, and Inference*. Taylor & Francis, 2010.

[8] O. W. Winkler, *Interpreting Economic and Social Data: A Foundation of Descriptive Statistics*, ser. Mathematics and Statistics. Springer, 2009.

[9] J. Ramsay and B. Silverman, *Functional Data Analysis*, ser. Springer Series in Statistics. Springer, 2005.

[10] W. Saeys, B. De Ketelaere, and P. Darius, "Potential applications of functional data analysis in chemometrics," *Journal of Chemometrics*, vol. 22, no. 5, pp. 335–344, 2008. [Online]. Available: http://dx.doi.org/10.1002/cem.1129

[11] J. Jacques and C. Preda, "Functional data clustering: A survey," *Advances in Data Analysis and Classification*, vol. 8, no. 3, pp. 231–255, 2014.

[12] X. Olive and P. Bieber, "Quantitative assessments of runway excursion precursors using mode s data," 06 2018.

[13] G. Jarry, D. Delahaye, F. Nicol, and E. Féron, "Aircraft atypical approach detection using functional principal component analysis," in *SESAR Innovations Days 2018*, 2018.

[14] G. Jarry, D. Delahaye, F. Nicol, and E. Feron, "Aircraft atypical approach detection using functional principal component analysis," *Journal of Air Transport Management*, vol. 84, p. 101787, 2020.

[15] A. Hassoumi, M. J. Lobo, G. Jarry, V. Peysakhovich, and C. Hurter, "Interactive shape based brushing technique for trail sets," 2019.

[16] N. Adrienko and G. Adrienko, "Spatial generalization and aggregation of massive movement data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 2, pp. 205–219, Feb 2011.

[17] R. Scheepens, N. Willems, H. van de Wetering, and J. J. van Wijk, "Interactive visualization of multivariate trajectory data with density maps," in *2011 IEEE Pacific Visualization Symposium*, March 2011, pp. 147–154.

[18] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '07. New York, NY, USA: ACM, 2007, pp. 330–339. [Online]. Available: http://doi.acm.org/10.1145/1281192.1281230

[19] P. Kalnis, N. Mamoulis, and S. Bakiras, "On discovering moving clusters in spatio-temporal data," in *Advances in Spatial and Temporal Databases*, C. Bauzer Medeiros, M. J. Egenhofer, and E. Bertino, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 364–381.

[20] D. Delahaye, S. Puechmorel, S. Alam, and E. Féron, "Trajectory mathematical distance applied to airspace major flows extraction," in *EIWAC 2017 The 5th ENRI International Workshop on ATM/CNS*, 2017.

[21] G. Jarry, N. Couellan, and D. Delahaye, "On the use of generative adversarial networks for aircraft trajectory generation and atypical approach detection," 2019.

[22] J. D. Cryer, *Time Series Analysis*, 1st ed. Belmont, CA, USA: Wadsworth Publ. Co., 1986.

[23] G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Kisilevich, and S. Wrobel, "A conceptual framework and taxonomy of techniques for analyzing movement," *Journal of Visual Languages & Computing*, vol. 22, no. 3, pp. 213–232, 2011.

[24] G. Andrienko and N. Andrienko, "A general framework for using aggregation in visual exploration of movement data," *The Cartographic Journal*, vol. 47, no. 1, pp. 22–40, 2010. [Online]. Available: https://doi.org/10.1179/000870409X12525737905042

[25] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale, "A descriptive framework for temporal data visualizations based on generalized space-time cubes," in *Computer Graphics Forum*, vol. 36, no. 6. Wiley Online Library, 2017, pp. 36–61.

[26] H. Hochheiser and B. Shneiderman, "Interactive exploration of time series data," in *Proceedings of the 4th International Conference on Discovery Science*, ser. DS '01. Berlin, Heidelberg: Springer-Verlag, 2001, pp. 441–446. [Online]. Available: http://dl.acm.org/citation.cfm?id=647858.738708

[27] P. Buono, A. Aris, C. Plaisant, A. Khella, and B. Shneiderman, "Interactive pattern search in time series," in *Visualization and Data Analysis*, 2005.

[28] R. Scheepens, C. Hurter, H. Van De Wetering, and J. J. Van Wijk, "Visualization, selection, and analysis of traffic flows," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 379–388, Jan 2016.

[29] C. Hurter, "Image-based visualization: Interactive multidimensional data exploration," *Synthesis Lectures on Visualization*, vol. 3, no. 2, pp. 1–127, 2015. [Online]. Available: https://doi.org/10.2200/S00688ED1V01Y201512VIS006

[30] C. Hurter, S. Conversy, D. Gianazza, and A. Telea, "Interactive image-based information visualization for aircraft trajectory analysis," *Transportation Research Part C: Emerging Technologies*, vol. 47, pp. 207–227, 2014.

[31] C. Hurter, B. Tissoires, and S. Conversy, "Fromdady: Spreading aircraft trajectories across views to support iterative queries," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1017–1024, Nov. 2009. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2009.145

[32] C. Hurter, N. H. Riche, S. M. Drucker, M. Cordeil, R. Alligier, and R. Vuillemot, "Fiberclay: Sculpting three dimensional trajectories to reveal structural insights," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 704–714, Jan 2019.

[33] R. A. Adams and J. J. Fournier, *Sobolev spaces*. Elsevier, 2003, vol. 140.

[34] H. Stark and J. Woods, *Probability, random processes, and estimation theory for engineers*. Prentice-Hall, 1986. [Online]. Available: https://books.google.fr/books?id=2pFRAAAAMAAJ

[35] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*. John Wiley & Sons, 2007, vol. 382.

[36] A. Hinneburg and D. A. Keim, "Optimal grid-clustering: Towards breaking the curse of dimensionality in high-dimensional clustering," 1999.

[37] P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, and S. Zeger, "Springer series in statistics," 2009.

[38] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013, pp. 160–172.